

EXHIBIT A

NOTE: “OR” operators have been omitted and “!” indicates a root expander.

1. (copyright! licens! author! publisher! book!) w/20 (infring! pira! violat! illegal! liab!

[REDACTED]

RFPs Set 4

- REQUEST FOR PRODUCTION NO. 77: Communications Concerning any licensing copyrighted works that were used to train the Meta Language Models.

RFPs Set 5

- REQUEST FOR PRODUCTION NO. 105: All Documents and Communications related to licensing the Llama Models and concerning the data used to train Llama Models, including but not limited to the inclusion of any copyrighted material within data used to train Llama Models.
- REQUEST FOR PRODUCTION NO. 106: All Documents and Communications related to Meta guidelines for including or excluding copyrighted material from data used to train Llama Models.

2. (copyright! cmi isbn) w/20 (delet! remov! strip! cut! strik! excis!)

RFPs Set 5

- REQUEST FOR PRODUCTION NO. 106: All Documents and Communications related to Meta guidelines for including or excluding copyrighted material from data used to train Llama Models
- REQUEST FOR PRODUCTION NO. 118: All Documents and Communications, including source code, relating to any efforts, attempts, or measures implemented by Meta to prevent Llama Models from emitting or outputting copyrighted material.
- REQUEST FOR PRODUCTION NO. 119: All Documents and Communications, including source code, relating to the processing of copyrighted material used in training Llama Models, including storage and deletion of copyrighted material.

3. “fair use” w/50 (genesis llama llm “large language model” train! book! dataset! “data set” database! emit! emis! output! [REDACTED])

RFPs Set 5

- REQUEST FOR PRODUCTION NO. 81: All Documents and Communications related to the decision to use Shadow Datasets for training Llama Models.
- REQUEST FOR PRODUCTION NO. 87: All Documents and Communications related to the decision to use books for coherent storytelling in the training of Llama Models.
- REQUEST FOR PRODUCTION NO. 117: All Documents and Communications relating to Meta’s decision to not identify databases used for training Llama Models.
- REQUEST FOR PRODUCTION NO. 123: All non-privileged Communications with third parties regarding Plaintiffs’ allegations and Meta’s actual or contemplated defenses, including fair use.
- REQUEST FOR PRODUCTION NO. 130: All Documents and Communications, including discussions, deliberations, or negotiations related to any actual, proposed, or contemplated licensing agreements (even if never written or executed) for AI training data, including any actual, proposed, or contemplated terms, conditions, and consideration.

4. (publish! w/5 book!) AND (licens! contract! agreement! negotiat! partner!)

RFPs Set 4

- REQUEST FOR PRODUCTION NO. 77: Communications Concerning any licensing copyrighted works that were used to train the Meta Language Models.

RFPs Set 5

- REQUEST FOR PRODUCTION NO. 130: All Documents and Communications, including discussions, deliberations, or negotiations related to any actual, proposed, or contemplated licensing agreements (even if never written or executed) for AI training data, including any actual, proposed, or contemplated terms, conditions, and consideration.
- 5. (“the eye” “the-eye.eu” “the pile” “thepile” sci-hub scihub sci-hub.pub) w/50 (book! database dataset “data set” train! llm llama “large language model”)

RFPs Set 5

- REQUEST FOR PRODUCTION NO. 81: All Documents and Communications related to the decision to use Shadow Datasets for training Llama Models.
- REQUEST FOR PRODUCTION NO. 83: All Documents and Communications related to the role of The Eye in the distribution of Books3.
- REQUEST FOR PRODUCTION NO. 87: All Documents and Communications related to the decision to use books for coherent storytelling in the training of Llama Models.
- REQUEST FOR PRODUCTION NO. 117: All Documents and Communications relating to Meta’s decision to not identify databases used for training Llama Models.
- REQUEST FOR PRODUCTION NO. 130: All Documents and Communications, including discussions, deliberations, or negotiations related to any actual, proposed, or contemplated licensing agreements (even if never written or executed) for AI training data, including any actual, proposed, or contemplated terms, conditions, and consideration.